116. **(Newly added).** A recombinant or isolated nucleic acid molecule which encodes glial cell line-derived neurotrophic factor having a molecular weight of about 31-42 kD on non-reducing SDS-PAGE, a molecular weight of about 20-23 kD on reducing SDS-PAGE, and which promotes dopamine uptake in dopaminergic neurons at a concentration of approximately 60 pg/ml.

Please cancel Claim 87 without prejudice. Applicants intend to pursue the antibody-related claims separately in a continuation application.

REMARKS

The relevant claim provisions have been amended to specify that the nucleic acid sequence may encode a protein having an amino acid sequence which at least approximately 70% identical to the amino acid sequence set forth in SEQ ID NO:4 or SEQ ID NO:6 when four gaps in a length of 100 amino acids may be introduced to assist in that alignment. The amendment provides a clearly defined meaning for the previously used "homology" terminology, and support for the amendment can be found in the specification at page 20, lines 12-22. The relevant claim provisions also have been amended to clarify the conditions for hybridization of the nucleic acid sequences. Support for this amendment can be found in the specification at page 63, lines 12-22 (Example 2). Thus, the claims have been amended to describe the disclosed sequences in terms of specific SEQ ID NOs, as well as structural and functional characterizations, as suggested by the Examiner.

The amendments should not be construed as an acquiescence to the rejections and have been made solely to expedite the prosecution of this application. Applicants reserve the right to pursue the claims as originally filed in another application(s). The amendments do not add any new matter. The above-listed claim amendments were made to better clarify the claims in response to the Examiner's objections, and thereby, make them suitable for allowance or place them in better form for appeal. Therefore, it is respectfully requested that the amendments be entered at this time.

Section 112, First Paragraph Rejections

The Examiner rejected claims 26, 31, 36, 42-55, 75-87 and 89-94 under §112, first paragraph, stating that the disclosure is enabling only for those claim provisions

108. **(Newly added)** A method for the expression of glial cell line-derived neurotrophic factor, comprising modifying a cell to express a nucleic acid sequence encoding a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:6 or a polypeptide which is at least 70% identical to the amino acid sequence set forth in SEQ ID NO:6 when four gaps in a length of 100 amino acids may be introduced to assist in that alignment.

109. **(Newly added)** The method of claim 108, wherein the expressed polypeptide is a monomer and wherein the method further comprises refolding expressed polypeptide to form a disulfide-bonded dimer.

110. **(Newly added)** The method of claim 109, wherein the disulfide-bonded dimer is glycosylated.

111. **(Newly added)** The method of claim 108, wherein the expressed polypeptide is secreted by said cell.

112. **(Newly added)** The method of claim 111, wherein the secreted polypeptide is a disulfide-bonded dimer.

113. **(Newly added)** A recombinant or isolated nucleic acid sequence, comprising:
(a) sequences comprising nucleotides -78 through -1 of SEQ ID NO:5, 59 through 208 of SEQ ID NO:8 or 59 through 289 of SEQ ID NO:25; or
(b) sequences encoding all or a portion of amino acids -26 to -1 of SEQ ID NO:5, 1 through 50 of SEQ ID NO:8 or -77 through -1 of SEQ ID NO:25;
wherein said sequences are used in the expression of glial cell line-derived neurotrophic factor.

114. **(Newly added)** A recombinant or isolated nucleic acid sequence comprising a sequence complementary to a nucleic acid sequence of claim 26

115. **(Newly added)** A recombinant or isolated nucleic acid sequence according to claim 26 further comprising an amino-terminal methionine residue.

which are limited to the sequences referred to by specific SEQ ID NOs. In particular, the Examiner stated that the specification does not provide one skilled in the art with a description of what is meant by the terms "homologous" and "hybridize". The Examiner, therefore, concluded that such nucleic acid sequences could not be obtained by those of ordinary skill in the art without resorting to an undue amount of experimentation. Applicants traverse this objection/rejection on the basis that the present disclosure, in accordance with §112, first paragraph, provides all of the information that is necessary for one of ordinary skill in the art to make and use these nucleic acid sequences without undue experimentation.

First, as suggested by the Examiner, the intended meaning of the term "homology" has been clarified, using the specific terminology of the specification. The description of percent homology or percent identity is clearly set forth in the specification by the definition of the means of calculating the percentage of homology. Applicants direct the Examiner's attention to page 20, lines 12-22, which states:

> The percentage of homology as described herein is calculated as the percentage of amino acid residues found in the smaller of the two sequences which align with identical amino acid residues in the sequence being compared when four gaps in a length of 100 amino acids may be introduced to assist in that alignment as set forth by Dayhoff, in *Atlas of Protein Sequence and Structure* Vol. 5, p. 124 (1972), National Biochemical Research Foundation, Washington, D.C., specifically incorporated herein by reference.

As described above, the claims have been amended to clarify that "homology" is defined by the percent of amino acids which are identical between the sequences being compared. Applicants, therefore, provided a clear description in the specification of how homology is determined, and with the amendment the claims are fully interpretable.

Second, the Examiner states that one skilled in the art would not understand how to make and use the nucleic acid sequences encoding homologous polypeptides. The Examiner states that one skilled in the art could not produce such a nucleic acid sequence without undue experimentation because there are no teachings in the specification or prior art teachings to enable one skilled in the art to rationally identify such sequences. In particular, the Examiner states that one skilled in the art could not produce a functional GDNF polypeptide in the form of a substitution variant without undue experimentation because there are no

prior art teachings to enable one skilled in the art to rationally design such a modified polypeptide. This aspect of the rejection is also respectfully traversed.

To fulfill the enablement requirement, an application must describe how to make and use the claimed invention. While experimentation needed to practice the invention must not be "undue experimentation", it is well known that enablement is not precluded by the necessity for some experimentation such as routine screening (see *In re Wands* 8USPQ2d 1400-1407 (CAFC, 1988)).

> "the determination of what constitutes undue experimentation in a given case requires the application of a standard of reasonableness, having due regard for the nature of the invention and the state of the art ... the test is not merely quantitative, since <u>a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed</u>." [emphasis added]

As demonstrated in the specification, following the initial discovery of a GDNF protein sequence, Applicants proceeded to identify rat and human nucleic acid sequences using the fully detailed procedures provided in Example 2. These hybridization procedures demonstrated that probes could readily be designed and used to identify various GDNF sequences across species. At the time of filing the present application it was also well known that naturally occurring sequences so identified could then be compared and that the conserved residues and regions in these sequences could readily be determined. Therefore, with the specific disclosure of the rat and human sequences, those skilled in the art were provided with the information needed to obtain naturally occurring variant nucleic acid sequences and identify conserved residues and regions in these sequences, just as was described in the specification. With this information, those skilled in the art were provided with the means of making and using nucleic acid sequences which encode the polypeptides as well as the blueprint for the rational design of non-naturally occurring variant proteins and their encoding nucleic acid sequences.

The specification describes precisely the means for identifying and isolating nucleic acid sequences encoding GDNF polypeptides. Using probes based on the first discovered protein, the rat nucleic acid sequence and polypeptide were identified (Please see Example 2(a)). Using probes based on the rat nucleic acid sequence, hybridization procedures (using conditions of reduced stringency) led to the identification of the human

nucleic acid sequence. (Please see Example 2(c)). Thus, the disclosure distinctly describes the means of obtaining and identifying variant sequences using defined and exemplary probes and hybridization techniques. It is also well within the competence of those of ordinary skill in the art to identify nucleic acid sequences from a variety of species using the sequences described in the specification. As a result, not only is the identification of the claimed sequences well within the competence of those of ordinary skill in the art, the specification also "provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed".

This guidance is also demonstrated with the further identification of nucleic acid sequences and amino acid sequences of other species such as chicken and Xenopus (South African clawed toad) attached hereto as Exhibit 1. Furthermore, it is well known that the polypeptides encoded by such sequences may readily be compared for percent identity, as described in the specification, to identify the amino acid residues and regions which are conserved between species, and conversely, those which vary. For example, the comparison of the rat and human amino acid sequences demonstrate a homology which can be described as an approximately 93% identical amino acid sequence.

With regards to the particular features, those skilled in the art readily appreciate that residues which are not conserved in nature might be modified. Moreover, as found in nature, such modifications are typically designed to be conservative amino acid substitutions so that the function of the polypeptide is not changed. These modifications are well with the knowledge and ability of the ordinarily skilled artisan. Evidence of the common availability and practice of these techniques is provided in the specification as well as in commonly used references as discussed further below. In addition to the disclosure of the specifically identified nucleic acid and protein sequences in the present specification, those skilled in the art were provided with the means as well as the guidance to identify, design, make and use variant proteins and the encoding nucleic acid sequences without the need for undue experimentation.

The ability to substitute one amino acid for another in a protein having a fully defined amino acid sequence is well established in the art. Once a starting sequence is known, it is well-appreciated that each of certain amino acid residues may be substituted with a different amino acid residue without affecting the structural integrity or activity of the molecule. For example, those of ordinary skill in the art appreciate that sequences having at least 70% identity and the preserved function of the native sequence involve modifications that are generally termed "conservative", *e.g.*, the substitution of one amino acid for another whose size, shape, charge, hydrogen bonding capacity or chemical reactivity are very similar. Such conservative variations have little or no effect on the overall net

charge, polarity or hydrophobicity of the variant molecule, and therefore, using these modifications there is a clear and reasonable expectation of successfully producing variant polypeptides having both the required homology and function.

The ability to select such amino acid residues for substitution is amply demonstrated by such common textbook descriptions of amino acid characteristics as found in Stryer's *Biochemistry*, 3rd Edition (pp. 16-21, W.H. Freeman and Company, New York (1988)). As taught in Stryer, the amino acids may be grouped based on their attributes, and the replacement or substitution of one member in a group with another member of that group will be conservative so as to have little or no effect on the overall net charge, polarity, or hydrophobicity of the protein. These conservative substitutions may be summarized as indicated below.

<u>Conservative amino acid substitutions</u>

<u>Basic</u>:	arginine
	lysine
	histidine
<u>Acidic</u>:	glutamic acid
	aspartic acid
<u>Polar</u>:	glutamine
	asparagine
<u>Hydrophobic</u>:	leucine
	isoleucine
	valine
<u>Aromatic</u>:	phenylalanine
	tryptophan
	tyrosine
<u>Small</u>:	glycine
	alanine
	serine
	threonine
	methionine

Furthermore, the skilled worker has ready access to commonly used molecular sequence analysis scoring systems. Such analysis systems are based on datasets similar to

that presented in Gribskov and Devereux, Sequence Analysis Primer (Chapter 3, pp. 134-137 and 233, Stockton Press, New York (1991)) which describes the work of Dayhoff *et al.* As depicted in the table on page 233, certain amino acid residues are known as being commonly substituted by nature. These naturally occurring substitutions generally correlate to the amino acid groupings discussed above in Stryer. For example: aspartic acid (D) may be replaced with glutamic acid (E) or asparagine (N); arginine (R) may be replaced with lysine (K); glutamine (Q) may be replaced with asparagine (N) or aspartic acid (D); and serine (S) may be replaced with threonine (T) or glycine (G). This widely accepted means of scoring protein sequence alignments is used as the standard by which the newer systems, such as BLAST and FASTA, compare themselves. Because such modifications occur naturally, a sequence modification plan made on the basis of such substitutions would not be expected to change the binding attributes or activity of the variant molecule. Thus, such conservative substitutions are clearly available for use to design variants using the isolated sequences as the protein blueprint.

In addition, recombinant technology and chemical synthesis procedures for the manufacture of such designed compounds are so well established as to be available in college textbooks and standard laboratory manuals. As cited in the application, one such common reference is Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (1989). This is a standard reference used by those skilled in the art to make modified nucleic acid sequences for the expression of variant polypeptides. For example, it provides commonly practiced procedures such as oligonucleotide-mediated mutagenesis which may be used to produce the variants (pp. 15.51-15.94).

When the present application was filed (*i.e.*, 1991), all of these analytical procedures and modification and production techniques were commonly known, routinely used and highly reproducible. The fact that this information appears in common textbooks and standard laboratory manuals further demonstrates that such information is neither unknown or obscure. Thus, this information is part of the common knowledge which those of ordinary skill in the art possess or know where to obtain. Furthermore, the standardization of these procedures and techniques exemplifies the fact that polypeptide and nucleic acid sequence variants can be designed and produced by known methods without undue experimentation. The novel GDNF nucleic acid and amino acid sequences, the means of determining homology, and the availability of these standard techniques, supports the conclusion that those skilled in the art were given sufficient guidance for the rational design of the claimed nucleic acid sequences once presented with the disclosure of the present specification.

Applicants agree that prior to the present disclosure of the specific GDNF polypeptides and nucleic acid sequences there was insufficient information to identify naturally occurring GDNF variants or to make variants by modifying the unknown sequences. With the discovery of those sequences and their disclosure in the specification, however, the level of ability in the art with reference to GDNF proteins was changed. Those of ordinary skill now had detailed descriptions of how to find, make and identify GDNF. In addition, they were provided with clear and concise information on the nucleic acid and amino acid sequences of GDNF. Such information, in combination with aspects of hybridization procedures and protein production that are well within the skill in the art, provides a broad scope of enablement.

It should also be noted that this aspect of the amended claims does not encompass all sequences, but rather only those GDNF-encoding sequences having specific structural and functional characteristics. For example, the provision in amended claim 26 is directed to nucleic acid sequences encoding polypeptides having at least approximately 70% identity, as discussed above, and having the ability to promote dopamine uptake in dopaminergic neurons. The fact that the claimed sequences encode a polypeptide having the function of human GDNF means the polypeptides retain the activity of that native protein. It does not require any procedure outside those described in the specification to make and use the claimed sequences or to evaluate this activity. Thus, the sequences are defined in terms of specifically identified sequences and function.

As demonstrated by the specification, not only was the step-by-step scheme of implementing the invention provided, it was successfully accomplished. Because there is sufficient guidance for one of ordinary skill to identify, rationally design, produce and test the presently claimed molecules following the teachings of the specification and the known art, it does not require undue experimentation to make and use the claimed subject matter, and therefore, the claims are sufficiently supported. The evidence comprising the teachings of the specification and the known art demonstrates that a skilled artisan would have been able to practice the full scope of the claimed subject matter. Because the specification enables the claimed subject matter, Applicants respectfully request that this rejection be withdrawn.

The Examiner also asserts that the rejection is applicable to the portions of claims using the term "hybridize". The Examiner states that the conditions for hybridization are not included in the specification.

Applicants traverse this rejection and respectfully direct the Examiner's attention to page 34, line 30, through page 35, line 20. The described hybridization conditions and

techniques, as well as the means of determining appropriate hybridization conditions, were well known to those of ordinary skill in the art at the time of filing the parent application. Beltz *et al.* (*Methods in Enzymology* 100:266-285, 1983), which is of record in this case, sets forth exemplary quantitative considerations as well as a common strategy for the systematic analysis of family members. In addition, textbook descriptions of hybridization techniques and conditions, such as Sambrook *et al.* (*Molecular Cloning, A Laboratory Manual*, 2nd edition) which was discussed above, provide descriptions of well-known, standard techniques. Further commonly used techniques for the selection of probes and hybridization procedures are described in Lathe, *J. Mol. Biol.*, 183, 1-12 (1985).

In addition to the disclosed general procedures, the specification clearly sets forth and fully describes several specific hybridization examples. In fact, both the rat and the human GDNF sequences identified in SEQ ID NOs 3 and 5 were identified using fully detailed hybridization techniques. Using a labeled degenerate sequence based on the novel protein purified in Example 1, a hybridization procedure was used to identify a nucleic acid sequence encoding a rat GDNF molecule (please see page 57, line 19 through page 58, line 26). Using a PCR-constructed probe made from the rat nucleic acid sequence (about 250 base pairs), a hybridization procedure was performed, using conditions of reduced stringency. This procedure identified a nucleic acid sequence, from a human genomic library, encoding a human GDNF molecule (please see page 62, line 4 through page 63, line 34). A similar procedure was used to identify human GDNF in a cDNA library constructed from A+ RNA extracted from the human putamen (please see page 64, line 7 through page 66, line 34). The disclosure of such reduced stringency procedures is clearly set forth in the present specification as well as in the Beltz *et al.* and Sambrook *et al.* references which are well known to those skilled in the art.

The claims, however, have been further amended to point out that conditions of reduced stringency are used to identify the sequences, as is described in Example 2. The sequences are further described as those which have the ability of promoting dopamine uptake in dopaminergic neurons. Whether or not a sequence has this combination of attributes is readily determined by the skilled practitioner, and therefore, the "metes and bounds" of the claims are readily understood. Moreover, the claimed subject matter is sufficiently enabled as the specification has provided specific examples and reasonable guidance to allow one of ordinary skill in the art to make and use the claimed subject matter. Because the Specification enables this aspect of the claimed subject matter, Applicants respectfully request that this rejection be withdrawn.

In summary, each of the presently claimed sequences are fully disclosed in the specification. The claimed sequences are delimited in relation to a specifically disclosed and referenced SEQ ID NO, a required amino acid sequence identity, and/or a required hybridization characteristic, as well as the requisite function of promoting dopamine uptake in dopaminergic neurons. Each of these requirements are clearly set forth in the specification with detailed descriptions of how to obtain the claimed subject matter. Thus, the claimed sequences are fully described in terms sufficient not only to envision but to clearly identify the sequences for those skilled in the art and to distinguish them from other materials. Furthermore, the claimed subject matter is enabled as the specification has clearly set forth specific examples and reasonable guidance to allow one of ordinary skill in the art to make and use the invention as presently claimed. Therefore, the specification fully enables the practice of the presently claimed subject matter, and the rejection for lack of enablement may properly be withdrawn.

Lastly, the Examiner rejected certain claims and claim provisions concerning the use of anti-GDNF antibody. These claims have been canceled, without prejudice, to be pursued in a copending application. Thus, the rejections may properly be withdrawn.

<u>Section 112, Second Paragraph Rejections</u>

Claims 26, 31, 42-55, 75-87 and 89-94 were rejected under §112, second paragraph. The Examiner restated that the use of the phrases "at least 70% homologous" and "at least 90% homologous" are indefinite as having no precise or accepted definition of the term "homologous".

This rejection is respectfully traversed. As discussed above, the specification clearly and precisely provide the definition for determining "homology". Because the present claims particularly point out and distinctly claim the subject matter of the present invention, it is respectfully requested that these rejections properly be withdrawn.

With respect to how the calculation is made, the Examiner provided an example of two sequences ABCDEF and ABEF. Using the definition provided in the specification, these sequences would align as follows:

```
AB--EF
||  ||
ABCDEF
```

where one gap in the length of the sequence was introduced to assist in that alignment. As a result, the percentage of homology or percent identity is 4/4 or 100%, *i.e.*, the optimal number of amino acid residues that identically align divided by the possible number of residues to be aligned. If the sequences were of equal length, using for example ABXYEF, the sequences are aligned without gaps as:

```
ABXYEF
||  ||
ABCDEF
```

and the percentage of homology is 4/6 or 67%.

Because Applicants have provided the definition to be used, there is no need to seek some other "definition available from teachings in the art" to understand and clearly delimit the claims. The description supporting the "homology" provisions of the pending claims has been provided in the specification, and the claims have been amended to clarify the metes and bounds of the claimed subject matter. Thus, this rejection may properly be withdrawn.

Claims 26 and 31 were objected to for improper "Markush" language. The claims have been amended and this objection is no longer applicable.

For the foregoing reasons and in view of the amendments, Applicants respectfully request reconsideration of and withdrawal of the outstanding rejections. Applicants' representative would appreciate the opportunity to talk with the Examiner, in person, to facilitate the prosecution of the application.

Respectfully submitted,

Daniel R. Curry

Daniel R. Curry
Attorney for Applicants
Registration No: 32,727
Phone: (805) 447-8102
Date: December 26, 1996

Please send all future correspondence to:

U.S. Patent Operations/DRC
M/S 10-1-B
AMGEN INC.
Amgen Center
1840 Dehavilland Drive
Thousand Oaks, California 91320-1789

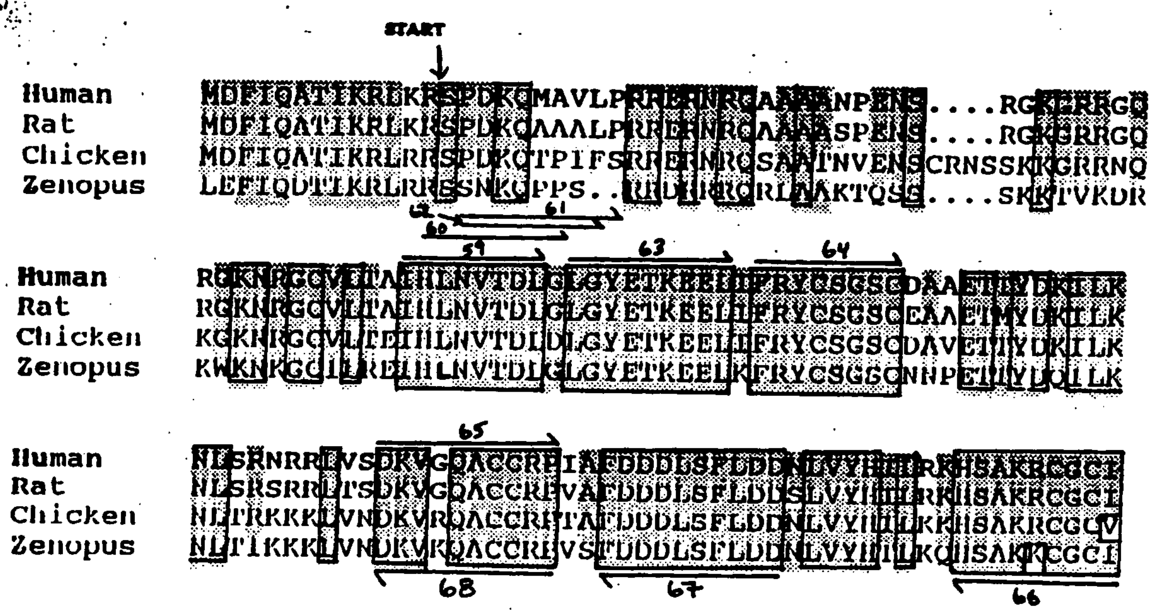
Exhibit I

From Page No.___

BASED ON CONSERVATION OF GDNF PHYLO-GENETICALLY - HUMAN, RAT, CHICKEN, XENOPUS

DESIGN NEW OLIGOS FOR PCR LIBS

BW RECENT XENOPUS SEQ + Hu, Ra, Ch :



ORDERED DEGENERATE OLIGOS + RI LINKER BASED ON SEQS AS MOST CONSERVED

OLIGOS DHD 59 → 68 SEQS ON p.29

To Page No. 29

Sequence Analysis Primer

Edited by

Michael Gribskov
and
John Devereux

M
stockton press

New York London Tokyo Melbourne Hong Kong

Extensions

The primary drawback to dynamic programming methods is that they require a considerable amount of computation. This limits their usefulness for tasks such as database searching. One simple way to speed up the alignment is to calculate only part of the score matrix, usually a diagonal band down the center (e.g., Sankoff and Kruskal, 1983). This can be safely done, for instance, if you know the sequences are homologous and do not require large gaps in their alignment, or if you have information from a faster method, such as hashing (see Hashing and Neighborhood Algorithms) that tells you where the most similar regions of the sequences are. Several methods that perform a banded alignment and iteratively increase the width of the band until the optimal alignment is found have been presented (Ukkonen, 1983).

A further great increase in alignment speed can be achieved through subdivision. If segments in each sequence can be identified, for instance by hashing methods (see Hashing and Neighborhood Algorithms), that are so similar they are unlikely to match with anything else, the alignment can be broken down into two smaller alignments, separated by the matching segment. Each equal subdivision increases the speed of the alignment by a factor of two.

Once a cDNA clone is sequenced, one usually wishes to identify the protein encoded by the message. One approach is to translate all three (or six) reading frames of the nucleic acid sequence and use the resulting protein sequences as probes in a fast database search (e.g., TFASTA - Lipman and Pearson, 1985; TBLASTN - Gish et al., in preparation). Unfortunately, this approach can be quite sensitive to frameshift errors in the cDNA sequence. An alternative to this approach (States and Botstein, 1990) uses dynamic programming methods to align the DNA and protein sequence.

SCORING SYSTEMS

The simplest scoring systems for molecular sequence analysis give positive scores only to comparisons of identical bases or residues. These scoring tables are referred to as an identity or unitary matrices and are still the primary scoring systems used for nucleic acid sequences.

The average rate of transition (purine to purine or pyrimidine to pyrimidine) mutations is about three times the average rate of transversion (purine to pyrimidine and vice versa) mutations. The rates of insertion/deletion mutations can also be determined from known homologous sequences. These values have been used to calculate scoring tables for nucleic acid sequences based on maximum likelihood methods (Bishop and Thompson, 1986). However, mutation rates and characteristics vary dramatically from species to species, from coding to non-coding regions, and from gene to gene, making it impossible to define a single best scoring system by

although it has been argued that it may be better to use an equivalent log-odds matrix calculated for a lower PAM for alignments of unknown sequences (Altschul, personal communication). The log of the probability of two sequences being evolutionarily related can, in principle, be calculated as the sum of the scores for each aligned pair of residues, i.e., the alignment score if a log-odds matrix is used as the scoring system for the alignment. However, this is overly simplistic since it ignores the effect of insertions and deletions on the probability.

The accepted point mutation model of protein sequence evolution is known to be imperfect in a number of ways. One common criticism is that the frequencies of mutations that require more than one base change in the DNA sequence is higher than would be expected for a simple Markovian model of DNA sequence evolution (Dayhoff and Eck, 1968; Wilbur, 1986). This criticism, however, is based on a specific model of DNA sequence evolution which is, itself, open to criticism (George et al., 1990). More importantly, the accepted point mutation model assumes that all residues in a protein are equally mutable; an assumption that is clearly incorrect. This can be easily seen by examining alignments of families of homologous proteins. For a set of six proteins, each sharing a pairwise sequence identity of 35 % or less, one would expect to find not a single amino acid conserved in every sequence if all positions were equally mutable. In actual families of proteins, however, it is not unusual to find several residues that are absolutely conserved in dozens of distantly related sequences (the active site triad of the serine proteases, for example). Furthermore, by starting from aligned sequences with only one or two differences, Dayhoff and coworkers selected mutations occurring at the most mutable positions as the basis for their calculation. Since the most important features in alignments are those positions that are unusually conserved, it has been argued that scoring systems based on the chemical or structural properties of the amino acid residues may produce better alignments (for example, Kubota et al., 1981; Feng and Doolittle, 1985; Argos, 1987; Risler et al., 1988). Lastly, the matrix may be biased because it was derived using mainly small globular proteins as a basis. In spite of its flaws, the MDM_{78} table, or a scoring table derived from it, remains the only widely accepted means of scoring protein sequence alignments. The important position that the MDM_{78} table occupies in molecular sequence analysis is clearly indicated by the fact that every new scoring system compares itself to the MDM_{78} table as a standard.

A special class of scoring systems known as metric distances merits additional consideration. A metric is a distance measure that has the following properties:

- no distance is less than 0
- identical sequence characters have a distance of 0
- the distance is symmetric; that is, the distance from A to B is the same as the distance from B to A

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
A	2	~~0~~	~~-2~~	0	0	-4	1	-1	-1	-1	-2	-1	0	1	0	-2	1	1	0	-6	-1	-3	0
B	0	0	-4	3	2	-5	0	1	-2	0	-3	-2	2	-1	1	0	0	0	-2	-5	0	-3	0
C	-2	-4	12	-5	-5	-4	-3	-3	-2	-5	-6	-5	-4	-3	-5	-4	0	-2	-2	-8	-3	0	-2
D	0	3	-5	4	3	-6	1	1	-2	0	-4	-3	2	-1	2	-1	0	0	-2	-7	-1	-4	1
E	0	2	-5	3	4	-5	0	1	-2	0	-3	-2	1	-1	2	-1	0	0	-2	-7	-1	-4	1
F	-4	-5	-4	-6	-5	9	-5	-2	1	-5	2	0	-4	-5	-5	-4	-3	-3	-1	0	-2	7	-2
G	1	0	-3	1	0	-5	5	-2	-3	-2	-4	-3	0	-1	-1	-3	1	0	-1	-7	-2	-5	0
H	-1	1	-3	1	1	-2	-2	6	-2	0	-2	-2	2	0	3	2	-1	-1	-2	-3	0	0	0
I	-1	-2	-2	-2	-2	1	-3	-2	5	-2	2	2	-2	-2	-2	-2	-1	0	4	-5	-1	-1	-1
K	-1	0	-5	0	0	-5	-2	0	-2	5	-3	0	1	-1	1	3	0	0	-2	-3	-1	-4	0
L	-2	-3	-6	-4	-3	2	-4	-2	2	-3	6	4	-3	-3	-2	-3	-3	-2	2	-2	-1	-1	-1
M	-1	-2	-5	-3	-2	0	-3	-2	2	0	4	6	-2	-2	-1	0	-2	-1	2	-4	-1	-2	-1
N	0	2	-4	2	1	-4	0	2	-2	1	-3	-2	2	-1	1	0	1	0	-2	-4	-1	-2	1
P	1	-1	-3	-1	-1	-5	-1	0	-2	-1	-3	-2	-1	6	0	0	1	0	-1	-6	-1	-5	0
Q	0	1	-5	2	2	-5	-1	3	-2	1	-2	-1	1	0	4	1	-1	-1	-2	-5	-1	-4	0
R	-2	0	-4	-1	-1	-4	-3	2	-2	3	-3	0	0	0	1	6	0	-1	-2	2	-1	-4	0
S	1	0	0	0	0	-3	1	-1	-1	0	-3	-2	1	1	-1	0	2	1	-1	-2	-1	-3	0
T	1	0	-2	0	0	-3	0	-1	0	0	-2	-1	0	0	-1	-1	1	3	0	-5	-1	-3	0
V	0	-2	-2	-2	-2	-1	-1	-2	4	-2	2	2	-2	-1	-2	-2	-1	0	4	-6	-1	-2	-1
W	-6	-5	-8	-7	-7	0	-7	-3	-5	-3	-2	-4	-4	-6	-5	2	-2	-5	-6	17	-3	0	-2
X	-1	0	-3	-1	-1	-2	-2	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-3	0	-2	0
Y	-3	-3	0	-4	-4	7	-5	0	-1	-4	-1	-2	-2	-5	-4	-4	-3	-3	-2	0	-2	10	-1
Z	0	0	-2	1	1	-2	0	0	-1	0	-1	-1	1	0	0	0	0	0	-1	-2	0	-1	0

This table is the log-odds form of the mutational distance matrix at 250 PAM (percent accepted mutation) as calculated by Dayhoff and co-workers (Dayhoff, 1978). This scoring table is probably the most commonly used in protein sequence comparisons and is also known as the MDM_{78} table. When comparing two sequences, the value in the row corresponding to a residue in the first sequence and the column corresponding to a residue in the second sequence indicates how likely these residues are to have arisen from unrelated sequences. Specifically, the values are the log of the probability that the residues resulted from mutation of a common ancestor, divided by the probability that they are related by chance. Positive values therefore indicate residues that are more likely than chance to have a common ancestor, and negative values indicate that an evoloutionary relationship is less likely than chance.

The PAM matrices are derived using a model of evolution wherein all positions are equally mutable, and are based on a specific set of observations of mutational frequency. For more details on the calculation of PAM matrices and their limitations see chapter 3 (Scoring Systems).

Values for B and Z are the averages of values for D and N, and E and Q, respectively. X is the average value for all comparisons.

BIOCHEMISTRY

THIRD EDITION

LUBERT STRYER
STANFORD UNIVERSITY

W. H. FREEMAN AND COMPANY / NEW YORK

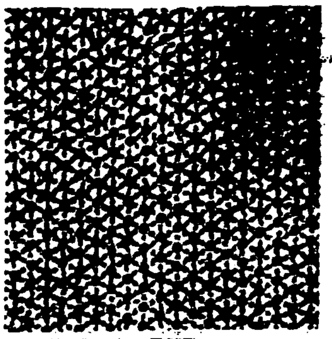


Figure 2-2
Electron micrograph of a cross section of insect flight muscle showing a hexagonal array of two kinds of protein filaments. [Courtesy of Dr. Michael Reedy.]

Figure 2-3
Electron micrograph of a fiber of collagen. [Courtesy of Dr. Jerome Gross and Dr. Romaine Bruns.]

3. *Coordinated motion.* Proteins are the major component of muscle. Muscle contraction is accomplished by the sliding motion of two kinds of protein filaments. On the microscopic scale, such coordinated motions as the movement of chromosomes in mitosis and the propulsion of sperm by their flagella also are produced by contractile assemblies consisting of proteins.

4. *Mechanical support.* The high tensile strength of skin and bone is due to the presence of collagen, a fibrous protein.

5. *Immune protection.* Antibodies are highly specific proteins that recognize and combine with such foreign substances as viruses, bacteria, and cells from other organisms. Proteins thus play a vital role in distinguishing between self and nonself.

6. *Generation and transmission of nerve impulses.* The response of nerve cells to specific stimuli is mediated by receptor proteins. For example, rhodopsin is the photoreceptor protein in retinal rod cells. Receptor proteins that can be triggered by specific small molecules, such as acetylcholine, are responsible for transmitting nerve impulses at synapses—that is, at junctions between nerve cells.

7. *Control of growth and differentiation.* Controlled sequential expression of genetic information is essential for the orderly growth and differentiation of cells. Only a small fraction of the genome of a cell is expressed at any one time. In bacteria, repressor proteins are important control elements that silence specific segments of the DNA of a cell. In higher organisms, growth and differentiation are controlled by growth factor proteins. For example, nerve growth factor guides the formation of neural networks. The activities of different cells in multicellular organisms are coordinated by hormones. Many of them, such as insulin and thyroid-stimulating hormone, are proteins. Indeed, proteins serve in all cells as sensors that control the flow of energy and matter.

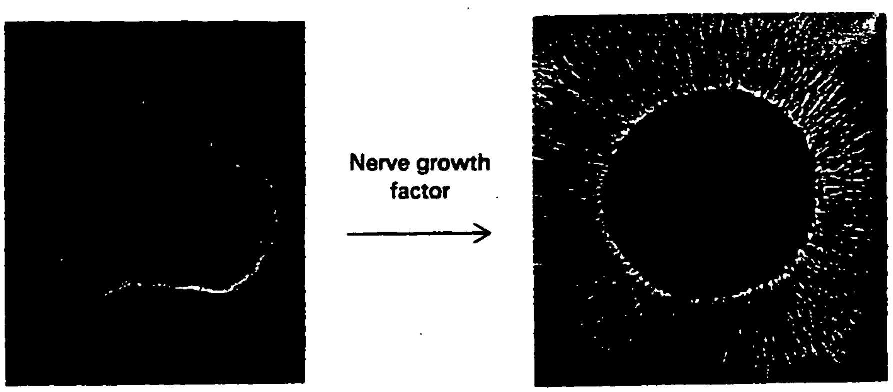


Figure 2-4
Photomicrograph of a ganglion showing the proliferation of nerves after addition of nerve growth factor, a complex of proteins. [Courtesy of Dr. Eric Shooter.]

PROTEINS ARE BUILT FROM A REPERTOIRE OF TWENTY AMINO ACIDS

Amino acids are the basic structural units of proteins. An α-amino acid consists of an amino group, a carboxyl group, a hydrogen atom, and a

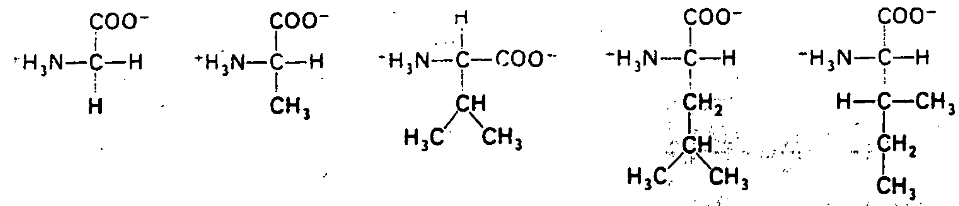


Figure 2-8
Amino acids having aliphatic side chains.

Let us look at this repertoire of amino acids. The simplest one is *glycine*, which has just a hydrogen atom as its side chain (Figure 2-8). *Alanine* comes next, with a methyl group as its side chain. Larger hydrocarbon side chains (three and four carbons long) are found in *valine*, *leucine*, and *isoleucine*. These larger aliphatic side chains are *hydrophobic*—that is, they have an aversion to water and like to cluster. As will be discussed later, the three-dimensional structure of water-soluble proteins is stabilized by the coming together of hydrophobic side chains to avoid contact with water. The different sizes and shapes of these hydrocarbon side chains (Figure 2-9) enable them to pack together to form compact structures with few holes.

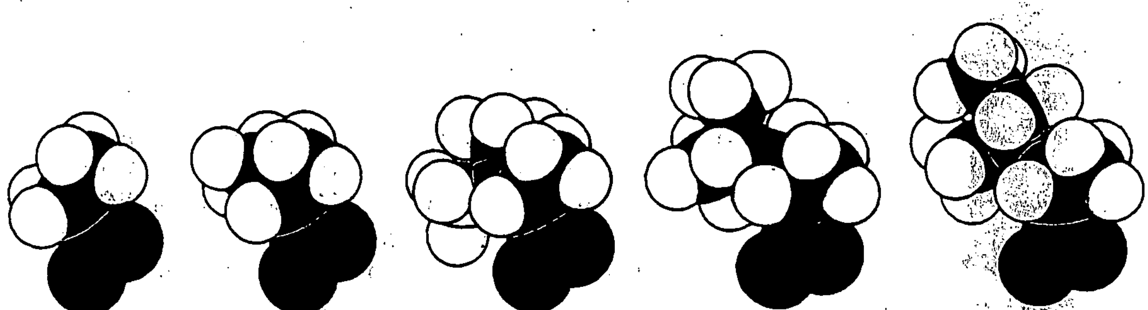


Figure 2-9
Models of aliphatic amino acids

Proline also has an aliphatic side chain but it differs from other members of the set of twenty in that its side chain is bonded to both the nitrogen and α-carbon atoms. The resulting cyclic structure (Figure 2-10) markedly influences protein architecture. Proline, often found in the bends of folded protein chains, is not averse to being exposed to water. Note that proline contains a secondary rather than a primary amino group, which makes it an *imino* acid.

Three amino acids with *aromatic side chains* are part of the fundamental repertoire (Figure 2-11). *Phenylalanine*, as its name indicates, contains a phenyl ring attached to a methylene (—CH_2—) group. *Tryptophan* has an indole ring joined to a methylene group; this side chain contains a nitrogen atom in addition to carbon and hydrogen atoms.

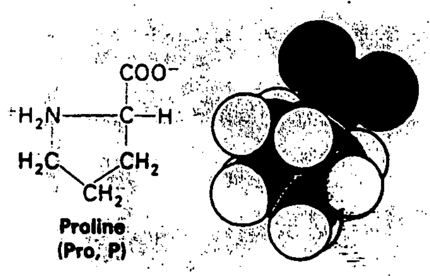


Figure 2-10
Proline differs from the other common amino acids in having a secondary amino group.

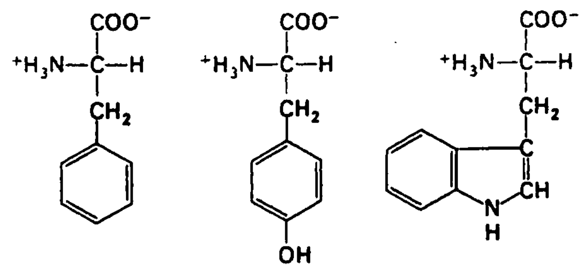
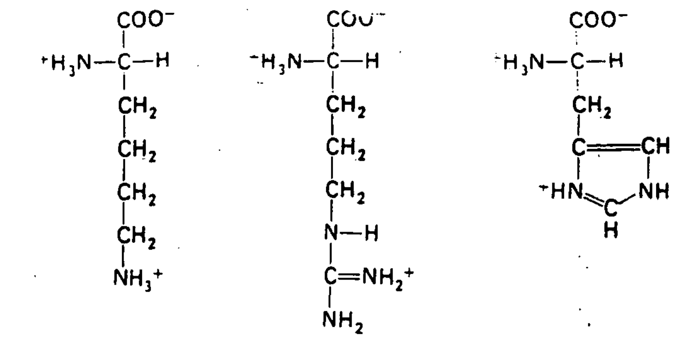


Figure 2-11
Phenylalanine, tyrosine, and tryptophan have aromatic side chains.

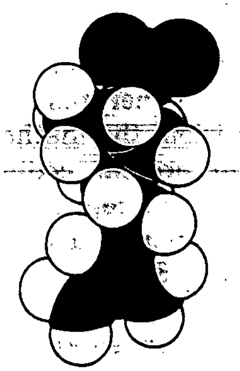


Lysine
(Lys, K)

Arginine
(Arg, R)

Histidine
(His, H)

Figure 2-16
Lysine, arginine, and histidine have basic side chains.

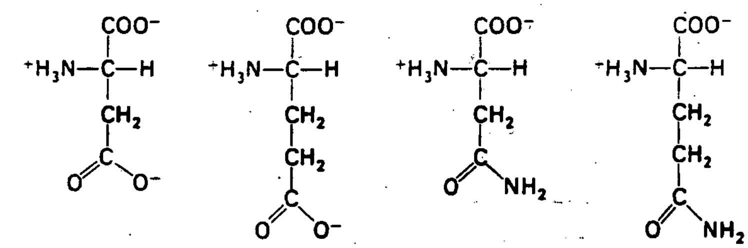


Arg

Figure 2-17
Model of arginine. The planar outer part of the side chain, consisting of three nitrogens bonded to a carbon atom, is called a guanidinium group.

sites of enzymes, where its imidazole ring can readily switch between these states to catalyze the making and breaking of bonds. These *basic amino acids* are depicted in Figure 2-16. The side chains of arginine and lysine are the longest ones in the set of twenty.

The repertoire of amino acids also contains two with *acidic side chains, aspartic acid* and *glutamic acid.* These amino acids are usually called *aspartate* and *glutamate* to emphasize that their side chains are nearly always negatively charged at physiological pH (Figure 2-18). Uncharged derivatives of glutamate and aspartate are *glutamine* and *asparagine,* which contain a terminal amide group in place of a carboxylate.



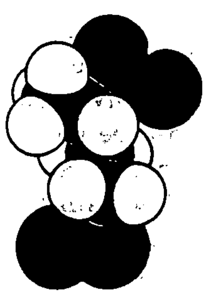
Aspartate
(Asp, D)

Glutamate
(Glu, E)

Asparagine
(Asn, N)

Glutamine
(Gln, Q)

Figure 2-18
Acidic amino acids (aspartate and glutamate) and their amide derivatives (asparagine and glutamine).



Glu

Figure 2-19
Model of glutamate.

Seven of the twenty amino acids have readily ionizable side chains. Equilibria and typical pK_a values for ionization of the side chains of arginine, lysine, histidine, aspartic and glutamic acids, cysteine, and tyrosine in proteins are given in Table 2-1. Two other groups in proteins, the terminal α-amino group and the terminal α-carboxyl group, can be ionized.

Amino acids are often designated by either a three-letter abbreviation or a one-letter symbol to facilitate concise communication (Table 2-2). The abbreviations for amino acids are the first three letters of their names, except for tryptophan (Trp), asparagine (Asn), glutamine (Gln), and isoleucine (Ile). The symbols for the small amino acids are the first letters of their names (e.g., G for glycine and L for leucine); the other symbols have been agreed upon by convention. These abbreviations and symbols are an integral part of the vocabulary of biochemists.

Oligonucleotide-mediated Mutagenesis

Oligonucleotide-mediated mutagenesis is used to add, delete, or substitute nucleotides in a segment of DNA whose sequence is known. In contrast to most other methods of mutagenesis, which typically spawn mixed populations of variants, oligonucleotide-mediated mutagenesis specifically generates mutations designed by the experimenter. Because of this precision, the method can be used, for example, to alter individual codons in protein coding sequences or to generate defined changes in sequences that have a regulatory function. In addition, oligonucleotide-mediated mutagenesis can be used to facilitate construction of new vectors and chimeric genes. For example, sequences such as ribosome-binding sites or polyadenylation signals can be inserted at predetermined positions in expression vectors; inconvenient restriction sites can be removed and convenient sites can be added at specified positions; and, finally, undesirable sequences (e.g., introns and DNA sequences that code for untranslated regions of mRNA) can be eliminated and different domains (e.g., promoters and coding regions) can be linked together with precision.

The feasibility of introducing specific changes at defined locations in DNA was first recognized in the early 1970s from work aimed at mapping the locations of mutations on the single-stranded genome of the small bacteriophage ϕX174. When fragments of denatured wild-type bacteriophage DNA were transfected into susceptible bacteria together with intact single-stranded bacteriophage DNA carrying an amber mutation, "marker rescue" was observed, i.e., bacteriophages carrying wild-type genomes were generated. Marker rescue occurred in the transfected bacteria because the fragment of wild-type DNA annealed to the corresponding sequence of the amber mutant, forming a mismatched heteroduplex that was converted by host-specified mismatch-repair systems into a full-length, wild-type genome. It was quickly realized that this process could also be used in reverse, i.e., that specific mutations could be introduced into wild-type DNA using mutated double-stranded fragments of viral DNA (Weisbeek and Van de Pol 1970; Hutchison and Edgell 1971). Later, when pioneering work in DNA chemistry had led to the routine synthesis of oligonucleotides (Letsinger and Lunsford 1976; Khorana 1979; Matteucci and Caruthers 1981), and when the availability and quality of DNA-modifying enzymes (DNA polymerase and DNA ligase) had improved, in vitro techniques for oligonucleotide-mediated DNA mutagenesis were developed. The first methods used synthetic oligonucleotides that were completely homologous to single-stranded bacteriophage ϕX174 DNA except for a single base change that, if incorporated into the bacteriophage genome, would generate a selectable phenotype. The oligonucleotides were annealed to single-stranded bacteriophage ϕX174 DNA and used as primers for DNA synthesis catalyzed in vitro by the Klenow fragment of *E. coli* DNA polymerase I. When the resulting heteroduplexes were transfected into bacteria, a dramatic increase was observed in the frequency of bacteriophages displaying the desired phenotype (Hutchison et al. 1978; Razin et al. 1978).

Preparation of Single-stranded Target DNA

All oligonucleotide-mediated mutagenesis procedures require a target DNA that is at least partially single-stranded. This can be prepared simply and efficiently by cloning the target DNA into bacteriophage M13 or into recombinant plasmids (phagemids) containing origins of replication derived from single-stranded bacteriophages (see Chapter 4). There are good reasons why the segment of target DNA cloned into these vectors should be as small as possible:

- Large segments of DNA can be unstable in single-stranded bacteriophage vectors and are prone to suffer spontaneous deletion.
- The chance that the mutagenic oligonucleotide will hybridize to an inappropriate site rather than to a desired sequence increases as the size of the target DNA increases.
- To ensure that the mutagenized target DNA contains only the desired mutation and no other, it is essential to sequence the entire fragment after oligonucleotide-mediated mutagenesis has been completed. The shorter the target DNA, the easier the task of determining its entire sequence.

In most instances, naturally occurring restriction sites can be used to insert an appropriately sized segment of target DNA (< 1 kb) into a single-stranded DNA vector. In a few cases, however, when no suitable restriction sites are present, it may be necessary to consider other options. These include:

- Cloning a fragment of target DNA that is larger than optimal.
- Carrying out preliminary oligonucleotide-mediated mutagenesis to introduce restriction site(s) at appropriate locations in the target DNA. This is worthwhile when many oligonucleotide-mediated mutations are to be introduced into the same region of DNA, for example, when generating a comprehensive set of substitutions of a particular amino acid. Under such circumstances, it is usually possible to take advantage of the degeneracy of the genetic code to introduce novel restriction site(s) upstream of or downstream from the target area without altering the sequence of amino acids for which the DNA codes. The resulting "cassette" of DNA can then be easily shuttled in and out of the bacteriophage M13 vector used for oligonucleotide-mediated mutagenesis.
- Avoiding bacteriophage M13 vectors altogether. Mutagenesis is then carried out using double-stranded DNA derived from a plasmid. A number of methods have been developed to avoid the use of single-stranded vectors, but all of them are comparatively inefficient and should only be used in desperation. These methods include: (1) exonucleolytic digestion of plasmid DNA that has been nicked at a specific site (Wallace et al. 1981a; Efimov et al. 1985), (2) denaturation of supercoiled DNA (Schold et al. 1984), and (3) formation of heteroduplexes between two linear DNA fragments such that the resulting molecule is circular and has a single-strand gap that includes the target site for mutagenesis (Oostra et al. 1983; Morinaga et al. 1984).

1985) or Sequenase™ (Schena 1989), which cannot readily remove the hybridized mutagenic primer from its template.

- The hybrid formed between the template and the 3′ terminus of the oligonucleotide is sufficiently stable to allow priming of DNA synthesis to occur with high efficiency. If the mismatched nucleotide is too close to the 3′ terminus, the 3′ segment of the oligonucleotide will be unable to form a stable hybrid with the target DNA and will therefore be susceptible to exonucleolytic degradation if the Klenow fragment of *E. coli* DNA polymerase I is used in the primer-extension reaction (Gillam and Smith 1979). In addition, an increase in the frequency of priming at incorrect locations might occur because the unhybridized 3′ region of the mutagenic oligonucleotide is now free to anneal to incorrect sites on the template. To suppress these effects, between 7 and 9 perfectly matched nucleotides are required at the 3′ terminus of the mutagenic oligonucleotide.

- The difference in thermal stability between perfectly matched hybrids and mismatched hybrids is sufficiently great that the mutagenic oligonucleotide can be used to screen bacteriophage M13 plaques by hybridization for potential mutants. As discussed in Chapter 11, the longer the oligonucleotide, the smaller the difference in thermal stability between a perfectly matched hybrid and one containing a single mismatched base pair. The aim, therefore, is to use the shortest mutagenic oligonucleotide that, under the conditions used for primer extension, will form stable hybrids both upstream of and downstream from the mismatch. Under normal circumstances, the mutagenic oligonucleotide should be between 17 and 19 nucleotides in length, with the mismatch centrally located.

2. *Oligonucleotides used to create deletions or insertions or to substitute two or more contiguous nucleotides.* Oligonucleotides 25 or more nucleotides in length are used to insert, delete, or substitute two or more bases. Optimally, there should be 12–15 perfectly matched nucleotides on either side of the central looped-out region to ensure that both ends of the mutagenic oligonucleotide are stably hybridized at the temperature used for primer extension. The thermal stability of each of the two flanking regions can be estimated from the following formula:

$$T_m = 4(G + C) + 2(A + T)$$

where T_m = melting temperature (in 6× SSC), G = number of G residues in the sequence, C = number of C residues in the sequence, A = number of A residues in the sequence, and T = number of T residues in the sequence. Mutagenesis occurs efficiently when the T_m of each of the double-stranded regions flanking the mismatched or looped-out sequence is approximately 35–40°C.

In general, the larger the size of the mutation to be constructed, the lower the efficiency of oligonucleotide-mediated mutagenesis. This inefficiency stems from two sources. First, the ability of the mutagenic oligonucleotide to form stable hybrids with two separate sequences on the

Hybridization of Oligonucleotides to the Template DNA and Primer Extension

In the standard double-primer method (Norris et al. 1983; Zoller and Smith 1984, 1987) (see Figure 15.7), two oligonucleotides—the phosphorylated mutagenic oligonucleotide and a universal sequencing primer (which need not be phosphorylated)—are mixed in a 10- to 50-fold molar excess with the single-stranded template DNA in a small volume of buffer. The template and oligonucleotides are heated briefly to 20°C above the estimated T_m, to denature any regions of secondary structure, and then cooled slowly to room temperature. Hybrids form as the temperature of the reaction mixture drops below the relevant T_m. The stoichiometry of the reagents in the mixture ensures that virtually all of the single-stranded template DNA is driven into hybrids with both oligonucleotides.

This protocol works well for mutagenic oligonucleotides of a wide variety of lengths and base compositions. However, annealing mixtures containing mutagenic oligonucleotides that are exceptionally rich in A + T may need to be cooled to lower temperatures (12–16°C) in order to stabilize the hybrids.

After the annealing reaction is complete, a mixture containing a DNA polymerase, dNTPs, DNA ligase, and ATP is added, and primer extension is allowed to proceed for 2–15 hours at the appropriate temperature for the polymerase being used. DNA synthesis is initiated both at the 3′ terminus of the mutagenic oligonucleotide and at the 3′ terminus of the upstream universal sequencing primer. The sequencing primer is extended by DNA polymerase until the 3′ terminus of the newly synthesized strand encounters the 5′ terminus of the phosphorylated mutagenic oligonucleotide (see Figure 15.7). The two segments of DNA then become ligated by the action of bacteriophage T4 DNA ligase. This sealing of the phosphodiester bond protects the 5′ terminus of the mutagenic oligonucleotide from displacement by the newly growing strand and prevents 5′ exonucleolytic editing of mismatched nucleotides after the DNA is transfected into *E. coli*. The 3′ terminus of the mutagenic oligonucleotide is also extended, resulting in the formation of a mutant wild-type heteroduplex (see Figure 15.7) that is transfected into the *E. coli*.

Any of several different DNA polymerases may be used in the extension reaction. Until recently, most workers relied exclusively on the Klenow fragment of *E. coli* DNA polymerase I, which lacks 5′ exonucleolytic activity and is therefore incapable of degrading the mutagenic oligonucleotide. Although the efficiency of mutagenesis is reported to be higher when bacteriophage T4 DNA polymerase (Geisselsoder et al. 1987) or Sequenase (Schena 1989) is used, the Klenow fragment generally gives yields of mutants that are more than adequate, and this remains the enzyme of choice.

Transfection of E. coli and Screening for Mutants

After the primer-extension reaction is complete, the resulting mixture of double-stranded heteroduplex DNA is then transfected directly into an appropriate bacterial host. Most of the transfected cells release bacteriophage particles that carry a wild-type copy of the target fragment. However, a portion of the transfected cells generate particles whose genomes carry the desired mutation. Mutants can therefore be identified by screening plaques by hybridization, using the radiolabeled mutagenic oligonucleotide as a probe.

Earlier protocols (see, e.g., Zoller and Smith 1982, 1983), in which the extension reaction was primed by a single, phosphorylated, mutagenic oligonucleotide, called for the purification of covalently closed circular DNA before transfection. The aim of this step was to reduce the background of plaques containing wild-type target DNA. However, this time-consuming enrichment procedure is no longer necessary because of the improvement in efficiency of mutagenesis brought about by the use of double primers (see page 15.57). In addition, advances in the understanding of the hybridization properties of small oligonucleotides (reviewed in Chapter 11) have made it possible to screen rapidly and simultaneously many thousands of plaques to identify those generated by bacteriophages carrying the desired mutation. Hybridization is usually carried out under conditions that allow the radiolabeled oligonucleotide to form hybrids with both mutant and wild-type DNA. By progressively increasing the stringency of the subsequent washes, it is almost always possible to find conditions that (1) cause dissociation of mismatched hybrids, such as those formed between the mutagenic oligonucleotide and wild-type DNA, and (2) do not dissociate perfect hybrids formed by the oligonucleotide and the desired mutant.

Methods to Improve the Efficiency of Oligonucleotide-mediated Mutagenesis

The basic procedures described above have been used successfully for several years to isolate a wide variety of site-directed mutants. Occasionally, however, difficulties are encountered in obtaining a particular mutation by the standard procedure. These difficulties have several causes:

- *The nature of the mutation itself.* The larger and more complex the mutation, the lower the efficiency with which it will be generated. For example, large deletions (i.e., deletions of several hundred nucleotides) are generated with approximately 50-fold lower efficiency than mutations involving only local changes in sequence. This decrease almost certainly occurs because the mutagenic oligonucleotide has problems in forming and maintaining stable hybrids with widely separated tracts of target DNA.
- *The nature of the target sequences.* Regions of single-stranded target DNA with a high propensity to form stable secondary structures (hairpin loops, stem loops, etc.) are difficult to mutagenize, presumably because such structures reduce the efficiency of annealing of the mutagenic oligonucleotide. A related problem may arise when the target DNA consists of, or contains, repeated sequences. In these cases, the mutagenic oligonucleotide may be able to anneal to sequences present at more than one location in the target DNA and may generate additional mutations at these sites. Thus, only a fraction of the plaques that hybridize to the oligonucleotide probe may actually carry the desired mutation.
- *The nature of the vector.* Oligonucleotide-mediated mutagenesis is generally carried out with templates generated from recombinant M13 bacteriophages. However, single-stranded templates can also be obtained by superinfection of bacteria that have been transformed with plasmids (phagemids) carrying an origin of DNA replication derived from a single-stranded bacteriophage vector (see Chapter 4). Although the phagemid system bypasses two time-consuming steps—cloning the fragment of target DNA into a bacteriophage M13 vector and recovering it after mutagenesis—the overall efficiency of mutagenesis is reduced by approximately five- to tenfold. This inefficiency results largely from variation in yield of single-stranded DNA from one superinfection experiment to the next and from one phagemid clone to another. Although the phagemid system can generally be used to isolate simple mutants (containing point mutations and small deletions or insertions), bacteriophage M13 vectors are preferred when constructing more complicated mutants.

To improve the efficiency with which such recalcitrant mutants can be isolated, a large number of variations of site-directed mutagenesis have been described (see, e.g., Marmenout et al. 1984; Bauer et al. 1985; Carter et al. 1985; Smith 1985; Taylor et al. 1985; Carter 1987; Zoller and Smith 1987). The best of these methods, however, was developed by Kunkel (1985; Kunkel et al. 1987) (see page 15.74) and takes advantage of a strong biological selection that can be applied against the wild-type strand of DNA used as template in oligonucleotide-mediated, site-directed mutagenesis. Because of

OLIGONUCLEOTIDE-MEDIATED MUTAGENESIS BY THE DOUBLE-PRIMER METHOD

The following method is modified from Zoller and Smith (1987).

1. In preparation for mutagenesis, clone a small fragment of DNA carrying the target sequence into an appropriate bacteriophage M13 vector (usually bacteriophage M13mp18 or M13mp19). Prepare single-stranded template DNA from a plaque generated by the recombinant bacteriophage. Methods for cloning into bacteriophage M13 vectors and for preparation of single-stranded bacteriophage DNA are given in Chapter 4.

   The mutagenic oligonucleotide should be designed as described on pages 15.54–15.56 and should be complementary to the strand of the target DNA that is packaged in bacteriophage M13 particles [the (+) strand]. Before use in site-directed mutagenesis, the mutagenic oligonucleotide should be purified by Sep-Pak C_{18} column chromatography to remove salts and other impurities (see Chapter 11, page 11.39). However, it is generally not necessary to purify the oligonucleotide by polyacrylamide gel electrophoresis unless the oligonucleotide is more than 30 nucleotides in length or is to be used for "loop-in" or "loop-out" mutagenesis.

2. Phosphorylate the mutagenic oligonucleotide with bacteriophage T4 polynucleotide kinase. Mix:

mutagenic oligonucleotide	100–200 pmoles
H_2O to 16.5 μl	
10 × bacteriophage T4 polynucleotide kinase buffer	2 μl
10 mM ATP	1 μl
bacteriophage T4 polynucleotide kinase	4 units

   Incubate the reaction for 1 hour at 37°C, and then heat the reaction for 10 minutes at 68°C to inactivate the polynucleotide kinase.

   10 × Bacteriophage T4 polynucleotide kinase buffer

   0.5 M Tris · Cl (pH 7.6)
   0.1 M $MgCl_2$
   50 mM dithiothreitol
   1 mM spermidine HCl
   1 mM EDTA (pH 8.0)

3. Anneal the phosphorylated mutagenic oligonucleotide and nonphosphorylated universal sequencing primer to the single-stranded bacteriophage M13 DNA containing the target sequence. Mix:

single-stranded template DNA (~1 μg)	0.5 pmole
phosphorylated mutagenic oligonucleotide	10 pmoles
nonphosphorylated universal sequencing primer	10 pmoles
10 × PE1 buffer	1 μl
H_2O to 10 μl	

able to displace the mutagenic oligonucleotide from its template. Use dNTPs of the highest quality to minimize the possibility that contaminating dUTP might be incorporated into the newly synthesized strand of DNA. The concentrated dNTP solutions sold by Pharmacia have worked well in our hands.

When using bacteriophage T4 DNA polymerase or Sequenase, the extension reaction should be incubated for 5 minutes at 0°C, 5 minutes at room temperature, and then 2 hours at 37°C. The low temperature optimizes initiation of DNA synthesis from the 3′ terminus, and the subsequent incubation at 37°C improves the efficiency of the extension reaction. In addition, the concentration of each of the four dNTPs in the reaction should be increased to 500 μM when using these polymerases. This increases the efficiency of the extension reaction and suppresses the strong 3′ exonuclease activity of bacteriophage T4 DNA polymerase.

5. Add 10 μl of the ice-cold PE3 mixture to the reaction mixture containing single-stranded DNA and annealed oligonucleotides (step 3). Incubate the final reaction mixture for 6–15 hours at 16°C.

6. Transfect competent *E. coli* of an appropriate host strain (e.g., TG1) as follows:

   a. Make a series of dilutions of the reaction mixture (1:10, 1:100, and 1:500) in 10 mM Tris · Cl (pH 7.6).

   b. In Falcon 2059 tubes, precooled to 0°C, combine 1- and 5-μl aliquots of the undiluted reaction mixture and of each dilution of the reaction mixture with 200-μl aliquots of competent TG1 cells (prepared as described in Chapter 1, pages 1.76–1.81).

   c. Store the mixtures on ice for 30 minutes, and then transfer them for exactly 2 minutes to a water bath equilibrated at 42°C.

   d. Remove the transfected cultures from the water bath, and mix each of them with 100 μl of an overnight culture of TG1 cells.

   There is no need to add TG1 cells if freshly prepared, rather than frozen, competent cultures are used.

   e. Add 2.5 ml of 2× YT top agar (melted and cooled to 45°C) to each culture, and plate each mixture on a separate YT agar plate. Incubate the plates for 16 hours at 37°C to allow plaques to form.

If mutagenesis is carried out by the Kunkel method (see pages 15.74–15.79), 1-μl and 5-μl aliquots of the undiluted reaction mixture are used to transfect competent cultures of *E. coli* strain TG1.

If single-stranded DNA derived from a phagemid such as pUC118 or pUC119 is used as the template for mutagenesis, 1-μl and 5-μl aliquots of the undiluted reaction mixture are used to transform competent cultures of *E. coli* strain MV1184 (see Chapter 4, page 4.15, for a description of this strain and Chapter 4, pages 4.37–4.38, for a description of the transformation protocol). Plate 50-μl and 100-μl aliquots of each transformation mixture onto LB agar plates containing 50 μg/ml ampicillin. Ampicillin-resistant colonies should appear after 18–24 hours of incubation at 37°C. Screen the transformed colonies as described on pages 15.72–15.73.

7. Screen plaques (pages 15.68–15.71) or colonies (pages 15.72–15.73) by hybridization with a radiolabeled oligonucleotide probe to detect putative mutants.

c. Wrap the strip of polyethyleneimine-cellulose in Saran Wrap and autoradiograph, or cut the strip horizontally into thin (0.25-cm) sections and measure the amount of radioactivity in each section in a scintillation counter.

Oligonucleotides will remain at the origin, whereas ATP and inorganic phosphate will migrate in the same direction as the solvent (inorganic phosphate migrates slightly slower than the solvent front and ATP is approximately equidistant between the origin and the inorganic phosphate). Thus, the transfer of phosphate from [γ-^{32}P]ATP to the oligonucleotide will result in the appearance of radioactivity at the origin. By measuring the amount of radioactivity at the origin and on the total strip, the percentage of radiolabel transferred from [γ-^{32}P]ATP to the oligonucleotide can be calculated. The specific activity of the probe can then be determined on the basis of the molar quantities of oligonucleotide and [γ-^{32}P]ATP in the reaction. Under the conditions described above, approximately 50% of the radioactivity should be transferred to the oligonucleotide, resulting in a specific activity of approximately 2500 Ci/mmole.

The transfer of radiolabel to the oligonucleotide can also be monitored by adsorption to DE-81 filters. The oligonucleotide binds tightly to the positively charged filters, whereas unincorporated radiolabel is removed by repeated washing with a solution of sodium phosphate. See Appendix E for details.

4. *(Optional)* Remove unincorporated radiolabel from the oligonucleotide by one of the methods described in Chapter 11, pages 11.33–11.39.

Generally, this step is necessary only when background hybridization is a persistent problem. Under normal circumstances, the unpurified reaction mixture may be used as a probe.

Hybridization with the radiolabeled oligonucleotide should be performed at a temperature 5–10°C below the T_m estimated from the following formula:

$$T_m = 4(G + C) + 2(A + T)$$

4. At the end of the hybridization period, cut off a corner of the plastic bag and pour the hybridization solution into a disposable plastic tube. Seal the tube, and store the solution at −20°C until it is needed for rescreening positive plaques (steps 10 and 16).

5. Quickly transfer the filters to a tray containing 200–300 ml of 6× SSC at room temperature. Cover the tray with Saran Wrap, and place it on a rotating shaker for 15 minutes. Replace the washing fluid every 5 minutes.

6. Quickly transfer the filters to a piece of Saran Wrap stretched on the bench. *Do not allow the filters to dry.* Cover the filters with another piece of Saran Wrap. Fold the edges of the two pieces of Saran Wrap together to form a tight seal. Apply adhesive dot labels marked with radioactive ink to the outside of the package and establish an autoradiograph by exposing the package of filters to X-ray film for 1–2 hours at −70°C, using an intensifying screen (see Appendix E).

Radioactive ink is made by mixing a small amount of ^{32}P with waterproof black drawing ink. We find it convenient to make the ink in three grades: very hot (>2000 cps on a hand-held minimonitor), hot (>500 cps on a hand-held minimonitor), and cool (>50 cps on a hand-held minimonitor). Use a fiber-tip pen to apply ink of the desired hotness to the adhesive labels. Attach radioactive warning tape to the pen, and store it in an appropriate place.

7. Compare the pattern of hybridization with the distribution of plaques. At this stage, it is normal to find that virtually every plaque hybridizes to the probe. Typically, however, some plaques hybridize more strongly than others, and these often turn out to be those that carry the desired mutation.

8. Transfer the filters to a plastic box containing 100–200 ml of 6× SSC that has been prewarmed to T_m − 10°C. Agitate the filters in the solution for 2 minutes, and then transfer them to a piece of Saran Wrap as described in step 6. Establish another autoradiograph. At this stage, it is often possible to identify two types of plaques: those whose radioactive signal has decreased in intensity and those that show no change in intensity.

To minimize dissociation of perfect hybrids formed between the radioactive oligonucleotide and the mutagenized target sequence, do not wash the filters for more than 2 minutes.

9. Repeat the cycles of washing and autoradiography, increasing the temperature of the 6× SSC washing solution by 2–3°C in each cycle. The aim is to find a temperature that does not markedly affect perfect hybrids but causes dissociation of mismatched hybrids (such as those

14. Isolate bacteriophage M13 replicative form DNA from a culture infected with plaque-purified recombinant bacteriophages (step 11) that carry the desired mutation and show no other changes in sequence in the target region. Methods to isolate and purify bacteriophage M13 replicative form DNA are given in Chapter 4.

15. Recover the mutated target sequence by digestion of bacteriophage M13 replicative form DNA with the appropriate restriction enzyme(s) and preparative gel electrophoresis. Reclone the target DNA in the desired vector.

16. Using a number of different restriction enzymes, digest aliquots of either a recombinant that carries the original (unmutagenized) target sequence or the recombinant (constructed in step 15) that carries mutagenized target sequence. Separate the resulting fragments by gel electrophoresis, and transfer them to a solid support (e.g., nitrocellulose filter or nylon membrane) as described in Chapter 9. Carry out Southern hybridization at $T_m - 10°C$, using the ^{32}P-labeled mutagenic oligonucleotide as probe. Wash the filter under the discriminatory conditions established in step 9 and autoradiograph. The final autoradiograph should show hybridization only to the relevant fragments of the mutagenized target DNA.

step 6, page 15.65). Phagemid DNA is isolated from the pooled colonies and used to transform another batch of competent MV1184 cells. The resulting colonies, which will contain pure populations of either mutant or wild-type phagemids, are then screened by hybridization, using the mutagenic oligonucleotide as probe.

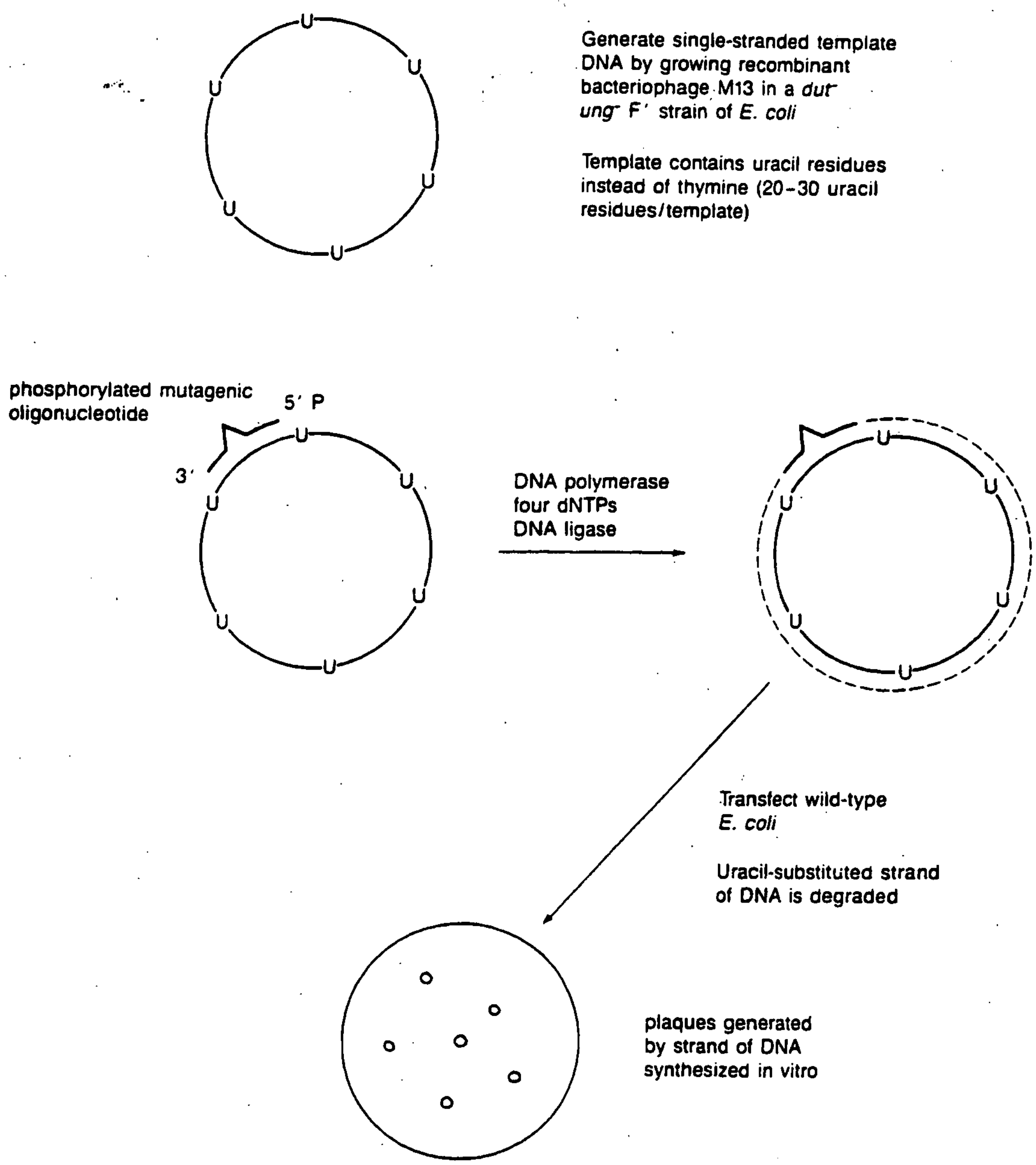


FIGURE 15.8
Oligonucleotide-mediated, site-directed mutagenesis using the Kunkel method (see text for details).

Chapter 4, there is a possibility that deleted variants will outgrow the original recombinant during extended periods of incubation. It is therefore advisable to verify that the majority of the single-stranded DNA used as template is of the correct size. Methods to analyze the size of bacteriophage M13 DNA by gel electrophoresis are described in Chapter 4, pages 4.39–4.40. It is also essential to sequence the entire segment of foreign DNA after mutagenesis to ensure that no deletions or other types of mutations have occurred at sites other than the immediate target sequence.

6. Measure the volume of the bacteriophage suspension, and then add 0.25 volume of NaCl/PEG solution. Mix the contents of the centrifuge bottle by swirling, and then store the bottle on ice for 1 hour.

   NaCl/PEG solution

   15% w/v polyethylene glycol (PEG 8000)
   2.5 M NaCl

7. Recover the precipitated bacteriophage particles by centrifugation at 5000*g* for 20 minutes at 4°C. Remove the supernatant by aspiration, and then invert the bottle to allow the last traces of supernatant to drain away. Use a pipette attached to a vacuum line to remove any drops of solution adhering to the walls of the bottle.

8. Resuspend the bacteriophage pellet in 4 ml of TE (pH 7.6). Transfer the suspension to a 15-ml Corex centrifuge tube, and wash the walls of the centrifuge bottle with another 2 ml of TE (pH 7.6). Transfer the washing to the Corex tube. Vortex the suspension vigorously for 30 seconds, and then store the tube on ice for 1 hour.

9. Vortex the suspension vigorously for 30 seconds, and then centrifuge it once more at 5000*g* for 20 minutes at 4°C in a fixed-angle rotor (e.g., Sorvall SS34).

10. Taking care not to disturb the pellet of bacterial debris, transfer the supernatant to a polypropylene tube. Extract the suspension twice with phenol equilibrated to pH 8.0 (see Appendix B) and once with phenol: chloroform. Separate the phases by centrifugation at 4000*g* for 5 minutes at room temperature. Avoid transferring material from the interface.

11. Transfer the aqueous phase from the final extraction to a glass centrifuge tube (e.g., a 30-ml Corex tube). Measure the volume of the solution, and add 0.1 volume of 3 M sodium acetate (pH 5.2), followed by 2 volumes of ethanol at 0°C. Mix the contents of the tube thoroughly, and then store the tube on ice for 30 minutes.

12. Recover the DNA by centrifugation at 5000*g* for 20 minutes at 4°C. Carefully remove the supernatant. Add 10 ml of 70% ethanol at room temperature, vortex briefly, and recentrifuge.

- Because growth of unmutated bacteriophages is suppressed with high efficiency, there is often no need to screen plaques by hybridization. Instead, single-stranded DNA can be prepared from a number of well-isolated plaques and analyzed directly by DNA sequencing.

Using Mutagenesis to Study Proteins

Until a few years ago, only two methods were available to study directly the relationship between the structure and function of a protein: (1) chemical modification of the side chains of the amino acids that form the primary sequence of the protein and (2) X-ray diffraction of protein crystals. Although both of these methods have yielded much information, they can be used only with proteins that are available in large quantity and are of high purity. A less direct type of analysis may be used if naturally occurring mutations can be identified that map in the gene coding for the protein of interest and that generate an observable phenotype. The mutant genes can then be cloned, and their DNA sequences can be compared to that of the wild-type allele. This approach has been invaluable in identifying mutations that affect the function of proteins such as the human LDL receptor (Davis et al. 1986) and factor VIII (Gitschier et al. 1988). However, this method also has its own set of restrictions: The mutations are sometimes confined to specific domains of the protein, and it is often difficult to isolate a sufficient number of independent mutants to allow definitive conclusions to be drawn about the structure of the protein.

These constraints can be partially circumvented by using site-directed mutagenesis to introduce mutations at predetermined sites in a cloned cDNA and then expressing the altered gene in an appropriate host-cell/vector system. By comparing the properties of the mutant and wild-type forms of the protein, it is often possible to identify domains or individual amino acid residues that are essential for the structural integrity and/or biological function of the protein. Because of the rapid advances that have occurred in recent years in both site-directed mutagenesis and expression of cloned genes, this method, which is sometimes called "reversed genetics," is often the first approach used by molecular biologists to analyze the relationship between a protein's structure and its function. The major problem with reversed genetics, however, is how to distinguish mutations that affect local structures from those that have profoundly deleterious effects on the folding or stability of the entire protein. Consider a typical experiment in which a number of point mutations have been generated at various sites in a gene coding for an enzyme. When the activities of these mutants are assayed, some of them show a reduction in catalytic function and others do not. In the absence of any other data, it is not possible to draw firm conclusions about the structure of the enzyme from this result. There is no way to know whether the substitution of one amino acid for another has affected only the structure and function of the active site or whether it has had more global effects. The problem would remain even if the three-dimensional shape of the wild-type enzyme were known. No algorithms have been yet devised that accurately predict the perturbations in protein structure caused by substitution, addition, or deletion of amino acid residues.

These difficulties can be alleviated by developing independent assays for the folding of the protein of interest. Such assays typically include:

2. Search the data banks for proteins with homologous amino acid sequences. Residues that are highly conserved between proteins of different function are more likely to be involved in forming common structural motifs than in highly specific and private activities such as catalysis or ligand binding.

3. Use a computer program to analyze whether the amino acid sequence contains internal repetitions. These indicate that the gene coding for the protein has evolved by one or more rounds of duplication followed by mutagenic drift. Residues that are conserved between different repetitive elements are more likely to be involved in forming common structural motifs than in catalysis or ligand binding.

4. When planning to make deletion mutants, avoid the temptation to remove the DNA sequences that lie between naturally occurring restriction sites. Although such deletions are easy to make, their borders are unlikely to lie at sensible positions within the coding sequence and the resulting proteins are almost certain to be malfolded.

5. If the positions of introns and exons have been established by analysis of the corresponding segment of genomic DNA, consider the possibility of creating a set of deletion mutants that lack particular exons or combinations of exons. This approach has been particularly useful in analyzing the structure and function of proteins that have evolved by exon shuffling (see, e.g., Gething et al. 1988). Because individual exons often encode independently folding polypeptide domains, there is a good chance that precise removal of a particular exon will not prejudice the folding of the remainder of the protein.

6. When constructing chimeras between nonhomologous proteins, try to arrange for the junctions between the two coding regions to lie at the borders of predicted structural domains rather than within them. For example, the ectodomains, transmembrane regions, or cytoplasmic tails of different transmembrane receptors should be exchanged at residues that are predicted to lie close to the appropriate transmembrane region. However, if the two proteins are homologous in sequence and function, the junctions should be located in segments of amino acid sequence that are identical or nearly so.

7. Avoid changing proline residues (all proteins) or cysteine residues (secretory or cell-surface proteins). These amino acids are usually involved in forming and maintaining essential structural motifs. Proline is the residue that is frequently used to terminate α-helical regions, whereas pairs of cysteine residues form the stabilizing disulfide bonds that are a hallmark of many secretory and cell-surface proteins.

8. When designing point mutations in secretory or transmembrane proteins, avoid altering asparagine or serine/threonine residues that lie within potential sites for *N*-linked glycosylation (Asn-X-Ser/Thr, where X is any amino acid other than proline). Mannose-rich oligosaccharides that are added posttranslationally to the asparagine residues in such consensus sequences are in some cases involved in assisting the nascent polypeptide to fold (Gallagher et al. 1988).

INSERTION OF HEXAMERIC LINKERS INTO PROTEIN-CODING SEQUENCES

The method given below, which is an elaboration of the work of Boeke (1981), is based on protocols published by Barany (1985a,b, 1987, 1988). Synthetic hexameric linkers are inserted quasi-randomly into the gene by targeting insertion to sites for frequently cutting restriction enzymes. The sequence of the hexameric linker is chosen so that no termination codons are inserted. A selectable marker (Vieira and Messing 1982) is used to identify mutants that have incorporated the linker. Although straightforward, the approach is limited to insertions at preexisting restriction enzyme sites. Nevertheless, this approach can be very useful in the initial determination of functional domains of a protein. The method involves the following steps:

- Cloning the target into a suitable plasmid (e.g., pUC18/pUC19 or pUC118/pUC119)
- Linearization of the plasmid DNA by partial digestion with a restriction enzyme that cleaves frequently within the target sequence and generates cohesive termini
- Ligation of a single-stranded hexameric linker to the cohesive termini
- Removal of excess linkers by cleavage with the appropriate restriction enzyme
- Ligation of a fragment of DNA carrying a selectable marker (usually *kan*ʳ) to the cohesive termini of the linkers
- Selection of transformed bacteria that are resistant to kanamycin and ampicillin
- Removal of the fragment of DNA carrying the selectable marker by cleavage with the appropriate restriction enzyme
- Recircularization of the linear DNA molecule lacking the *kan*ʳ gene
- Transformation of bacteria with the recircularized plasmid
- Screening of transformed colonies for those that carry plasmids containing a novel restriction site within the target sequence

This method is therefore a variation of the technique described on pages 15.32–15.50 that is used to construct linker-scanning mutations. A preassembled kit can be purchased from Pharmacia (TAB™ mutagenesis system) that contains the materials required for this type of mutagenesis.

```
        5′...GCGC...3′
        3′...CGCG...5′
              |
              | HhaI
              ↓
     5′...GCG      C...3′
     3′...C      GCG...5′
              |
              | 5′AATTCG3′
              ↓
5′...GCG        AATTCGC...3′
3′...CGCTTAA        GCG...5′
```

When the appropriate hexameric linker has been synthesized, proceed as described on pages 15.88–15.94.

bromide has been added, use 1 × restriction enzyme buffer to adjust the volume of the solution in each tube to 49 μl. The concentrations of ethidium bromide in the three digestion mixtures are therefore 20 μg/ml, 40 μg/ml, and 60 μg/ml, respectively.

Caution: Ethidium bromide is a powerful mutagen and is moderately toxic. Follow precautions detailed in step 1h, page 15.9.

b. Add 10 units of the appropriate restriction enzyme to each tube, and immediately transfer the three tubes to a water bath set at 37°C. After 0, 5, 10, 20, 30, 45, and 60 minutes of incubation, transfer aliquots (5 μl) from each of the three tubes to fresh microfuge tubes containing 15 μl of 5 mM EDTA (pH 8.0). Heat each sample for 10 minutes at 68°C. Store the tubes containing the aliquots on ice until all of the samples have been collected.

c. Analyze the DNA in each sample by electrophoresis through a 0.8% agarose gel cast and run in 0.5 × TBE (see Appendix B) containing ethidium bromide (0.5 μg/ml). As markers, use plasmid DNA that has been linearized by digestion with a restriction enzyme that cleaves at only one site. Under these conditions of electrophoresis, closed circular DNA migrates slightly faster than linear DNA and considerably faster than relaxed circular DNA.

d. Examine the gel by ultraviolet illumination, and determine the conditions that give the maximum yield of full-length linear molecules. Usually not more than 33% of the original closed circular plasmid DNA is converted to full-length linear molecules.

Caution: Ultraviolet radiation is dangerous, particularly to the eyes. Follow precautions detailed in step 1i, page 15.9.

e. Set up a large-scale digestion containing 20 μg of closed circular DNA. Increase the volume of all of the components of the reaction proportionally, and incubate the reaction for the time required to achieve maximal conversion of closed circular DNA to full-length linear molecules.

5. Purify the full-length linear DNA by preparative agarose gel electrophoresis using one of the methods described in Chapter 6. Redissolve the DNA in TE (pH 7.6) at a concentration of 250 μg/ml. Approximately 1–2 μg of full-length linear DNA will be required to complete the remainder of the procedure.

6. a. *Ligation of single-stranded hexameric linkers to linear plasmid DNA carrying protruding 5' termini and removal of excess linkers*

i. Phosphorylate approximately 10 μg of hexameric linkers with bacteriophage T4 polynucleotide kinase as described on page 15.63. The total volume of the reaction should not exceed 10 μl.

ii. Ligate the phosphorylated linkers to the linear plasmid DNA as follows:

iii. Add:

10× bacteriophage T4 polynucleotide kinase buffer	2 μl
10 mM ATP	2 μl
H_2O	6 μl
bacteriophage T4 polynucleotide kinase	10 units

Incubate the reaction at 37°C for 30 minutes, and then heat the mixture to 68°C for 15 minutes to inactivate the bacteriophage T4 polynucleotide kinase.

10× Bacteriophage T4 polynucleotide kinase buffer

0.5 M Tris·Cl (pH 7.6)
0.1 M $MgCl_2$
50 mM dithiothreitol
1 mM spermidine HCl
1 mM EDTA (pH 8.0)

iv. Purify the DNA by extraction with phenol:chloroform, and recover the DNA from the aqueous phase by precipitation with 0.5 volume of ammonium acetate and 2.5 volumes of ethanol. Store the tube for 10 minutes on ice, and recover the DNA by centrifugation at 12,000*g* for 15 minutes at 4°C in a microfuge. Remove the supernatant and wash the pellet of DNA carefully with 70% ethanol. Allow the pellet to dry at room temperature, and then redissolve it in 20 μl of TE (pH 7.6).

v. Proceed to step 7.

7. Add:

H_2O	18 μl
10× ligase buffer	5 μl
purified fragment of DNA that carries the *kan*ʳ gene (step 3) (0.5 μg)	2 μl

Add 5 Weiss units of bacteriophage T4 DNA ligase and 10 mM ATP to a final concentration of 1 mM. Incubate the reaction for 6–8 hours at 16°C.

In a separate tube, set up a control containing all of the components of the ligation mixture except the linearized plasmid DNA to which hexameric linkers have been added and incubate as above.

8. Use 10-μl aliquots of the control and test ligations to transform competent *E. coli* of an appropriate strain (e.g., DH1 or DH5) to resistance to kanamycin and ampicillin. Select transformants on LB agar plates containing kanamycin and ampicillin (or carbenicillin), each at a concentration of 100 μg/ml. If the experiment has gone well, the test ligation should generate approximately tenfold more colonies than the control ligation.

ampicillin and kanamycin, each at a concentration of 100 μg/ml. Use a bent glass rod and a pasteur pipette to scrape and squirt the colonies from the surfaces of the plates. Pool the bacterial suspension obtained from each of the plates, and recover the bacterial cells by centrifugation at 5000*g* for 10 minutes at 4°C. Remove the supernatant medium.

16. Isolate the closed circular plasmid DNA from the pooled colonies by one of the small-scale methods described in Chapter 1.

17. To remove the fragment carrying the *kan*^{r} gene from the plasmids containing the synthetic linker, digest the preparation of pooled plasmid DNAs with a restriction enzyme that recognizes the site created by ligation of the hexameric linkers.

18. When the restriction digest is complete, dilute the reaction mixture with 1× ligase buffer until the concentration of the plasmid DNA is <3 μg/ml. Add approximately 5 Weiss units of bacteriophage T4 DNA ligase per 100 μl of reaction mixture. Incubate the reaction for 4 hours at 16°C.

1× Ligase buffer

20 mM Tris · Cl (pH 7.6)
5 mM $MgCl_2$
5 mM dithiothreitol
1 mM ATP

19. Concentrate the DNA by extracting the ligation mixture twice with 1-butanol (see Appendix E). Add 0.2 volume of 10 M ammonium acetate, and precipitate the DNA with 2.5 volumes of ice-cold ethanol. Store the tube in an ice bath for 30 minutes.

20. Collect the DNA by centrifugation at 25,000*g* for 30 minutes at 0°C. Carefully decant the ethanol, and wash the pellet of DNA with several milliliters of 70% ethanol at room temperature. Again, recover the DNA by centrifugation, and remove the 70% ethanol by careful aspiration. Stand the open tube on the bench until the last traces of ethanol have evaporated. Dissolve the DNA in 20 μl of TE (pH 7.6), and estimate the recovery of DNA by analyzing a small aliquot by electrophoresis through an agarose gel.

21. Use approximately 50 ng of the DNA to transform competent *E. coli* (strain DH1 or DH5) to resistance to ampicillin *only*. Plate the transformants on LB agar medium containing ampicillin (100 μg/ml), and incubate them for 18–24 hours at 37°C. Store the remainders of the ligation mixtures at −20°C.

22. Pick 36 independent transformants, and establish small-scale (2-ml) cultures in LB medium containing ampicillin (100 μg/ml). After the